

Using the deepest regression method for optimization of fluidized bed granulation on semi full scale

1. Introduction

Experimental designs are widely used in the pharmaceutical sciences (Doornbos, 1981). Experimental design dealing with the granulation process has been applied in several studies (Dussert et al., 1995; Gordon, 1994; Gorodnichev et al., 1981; Lipps et al., 1994; Merkkü et al., 1994; Meshali et al., 1983; Miyamoto et al., 1995; Vojnovic et al., 1995). In all these studies the collected data were complete, which facilitated statistical analysis. However, sometimes the data is not complete and the usual statistical analysis is not applicable. Montgomery (1991) proposed to apply regression methods for modeling when incomplete data occurred in experimental design. In the previous study (Rambali et al., 2001b) regression models were developed for the granule size by omitting the missing data. However, the regression model could become biased due to the incomplete data and is valid only for the investigated domain. Therefore, complete data must be used for the development of a regression model in order to be valid for the whole experimental domain. Rousseeuw and Hubert (1999) proposed the deepest regression method to handle censored data. In the present study, the deepest regression method was used to estimate the missing granule size data. The objective of this study was to investigate the applicability of deepest regression on the incomplete data in the experimental design and subsequently develop regression models (fundamental and empirical) for the granule size based on the “complete data” (the observed data and the estimated missing data), which were valid for the whole experimental domain. These models will be evaluated for their adequacy and compared with the models developed in the previous study (Rambali et al., 2001b).

2. Materials and Methods

2.1 Granulation process

For the specification of the granulation process and the determination of the granule size we refer to the previous study (Rambali et al., 2001b). The calculation and the background theory of the theoretical powder bed moisture content and the ratio for the droplet size are also referred to this study.

2.2 Statistical analysis

For the experimental design development and the settings of the process variables we refer to the previous study (Rambali et al., 2001b). The settings of the process variables are listed in table 1. The design is displayed in table 2. Some runs did not succeed and the deepest regression method was used to estimate the granule size of these runs. In the next section this method is explained. Multiple regression modeling was used for optimization of the granule size. The optimum desired was between 300 μm and 500 μm .

The design was developed by the graphic software "STATGRAPHICS PLUS" version 3.3 (STSC Inc., Rockville, MD, USA). The statistical analyses were also carried out by this software. To compute the deepest regression we used the algorithm MEDSWEEP of Van Aelst et al., 2002 which is available as a stand-alone FORTRAN program and can be downloaded from the website <http://win-www.uia.ac.be/u/statis/>.

2.3 Deepest regression

Rousseeuw and Hubert (1999) defined the regression depth of a fit in a regression model. The regression depth of a regression fit measures how well the data are balanced about the model; hence the regression depth indicates how well the regression fit actually fits the data. The regression depth ranks all the possible regression fits for a given data set. Fits with a higher

regression depth better fit the data than do models with a lower regression depth. This immediately leads to the definition of the deepest regression estimator which is the model with maximal regression depth relative to the data. Let us consider an example to clarify the regression depth theory. A simple regression data set is given in figure 1. Three fits are depicted for the data set in the figure. Model M is the trendline of the data set. Models K and L do not fit the data as good as model M. The regression depth of a model indicates how well the model fits the data. Any possible fit which can be tilted in some way until it becomes vertical (such as lines *a* or *b* in figure 1) without passing (or touching) any observations is called a nonfit. Model K is a nonfit, because when it is tilted to become line *a*, it does not pass through an observation. Models L and M pass through one and two observations respectively, when they are tilted to become line *a*. The regression depth of a model is now defined as the smallest number of observation that has to be removed from the data in order to make a model a nonfit. In the example given in figure 1 the regression depth is respectively 0, 1 and 2 for the models K, L and M.

The deepest regression (DR) is a robust method compared with the least square method, when outliers do occur in the data set. To show the robustness of the DR, let us consider the breakdown value. The breakdown value of a model is the smallest fraction of the data set, that must be replaced by arbitrary values to make the method explode. The breakdown value for the DR is always at least $1/(p+1)$ (p is the number of the parameters in the model). In fact it converges to $1/3$ (when the number of points in the data set goes to infinity). The breakdown value for the least square is zero, which means that a single outlier in a data set can make the least square estimates completely useless.

Note from figure 1 that the regression depth only depends on the position of the observations in relation to the model; in other words it depends only on the sign of the residual corresponding to the estimate and on the x -values. This allows us to apply the deepest

regression to data with censored responses. If we do not know the actual response of an observation but we do know the sign of its residual, then we can still compute the deepest regression. Hence, we do not have to delete the observation from the data set. We simply set the missing responses equal to some arbitrary large value (relative to the measured responses) multiplied by the sign of the residual corresponding to the observation. If we now compute the deepest regression for a model, it follows that the residuals corresponding to observations with missing response have the desired sign. The exact method how this is done for the censored data in this study is described by Van Aelst et al, 2001. More discussion on deepest regression has been given by Bai and He, 1999, Rousseeuw and Hubert, 1999, Van Aelst and Rousseeuw, 2000 and Van Aelst et al., 2002.

3. Results and Discussion

3.1 Results

The results of the design are given in table 2.

Eight runs did not succeed, because the powder bed was overwetted. Visual inspection showed a wet powder bed with a slurry surface in these runs. In the previous study (Rambali et al., 2001b) the results of these runs were omitted in order to develop an empirical and a fundamental model for the granule size. However, these models were only valid for a part of the experimental domain. The question we asked was whether a model could be developed for the whole experimental domain, including the overwetted runs. In order to develop this model, the granule size of the missing runs should be estimated. Before the granule size of the unsuccessful runs can be estimated some indication about the granule must be obtained. The overwetting mechanism can give some indication about the granule size.

The overwetting mechanism of a fluid bed granulation is described by Schaafsma (2000). In the fluidized bed granulator, two different zones can be distinguished in which different stages

of the agglomeration process take place. The first zone is at the surface, the wetting zone, where the liquid droplets collide with the powder particles. In this zone the liquid concentration is high. Beneath the wetting zone and above the distribution grid the second zone is distinguished, where the wetted particles are mixed with the primary particles. In this zone the agglomerates are dried. In the wetting zone, the agglomerate growth depends on the spraying rate, the droplet size and the renewal rate of new particles (depending on the airflow rate). Depending on the airflow rate, the agglomerate can be dry or still (partially) wet when it reappears at the spray surface. This is important for the growth rate. When the agglomerates are dry the particle surface will absorb the binder liquid when it is wetted. When the particle is still wet, less liquid is absorbed and more liquid is available for further growth of the agglomerate. At large liquid concentration, the granule growth could become uncontrollable. Looking at the granule size data, some indication for the granule size of the missing runs can be observed. The results in table 2 indicate that all the missing runs have high powder bed moisture content after the spraying cycle and low level of inlet airflow rate and inlet air temperature and/or high level of spray rate in their variable combination. These observations confirmed that probably an uncontrollable granule growth occurred and resulted in defluidization of the powder bed. In order to fluidize the powder bed, the airflow must be high, as it was the case in run 12. The granule size of run 12 indicated that at overwetting, with high powder bed moisture content, an uncontrollable granule growth occurred which resulted in large granules. Because the overwetted powderbed became defluidized in the unsuccessful runs, the granule size could not be determined. The airflow rate settings of those overwetted runs, necessary for fluidizing the powder bed and overcoming the cohesive force between the granules, were lower than in run 12. Based on the settings and the result of run 12, it was expected that the granule size of the overwetted runs was larger than that of run 12. Therefore the granule size of these runs in table 2 are indicated “> 811 μm ”. Setting this

threshold for the unsuccessful runs, it allows to estimate the granule size using the deepest regression method developed by Rousseeuw and Hubert (1999). This method requires a model for the granule size. In the previous study (Rambali et al., 2001b), two models were proposed for the granule size: an empirical model and a fundamental model. The empirical model was based on the process variables and the fundamental model was based on the granule growth factors such as the powder bed moisture after the spraying cycle, a measure for the droplet size and the deformation force exercised by the airflow rate.

3.2 Empirical model

In the previous study (Rambali et al., 2001b) a quadratic model for the granule size was proposed. A deepest regression was applied on this model. The estimated granule size for the overwetted runs by deepest regression for the empirical model is shown in table 2. Note that the estimated granule size for the overwetted runs was larger than 811 μm (run 12) and was therefore compatible with what was observed. As expected, larger granule sizes were estimated for the runs with the most unfavorable granulation process setting, at low inlet air temperature and airflow rate and at high spray rate (runs 5 and 24). As discussed above, due to overwetting, the powder bed defluidized and the granule size could not be determined. The estimated granule size correctly indicated that at such process conditions, the granulation will not be optimal and therefore the model based on the estimated data is useful to avoid such unfavorable process settings.

Based on the completed data, a stepwise multiple regression was applied, which means that non-significant parameters were sequentially eliminated and only significant coefficients ($P < 0.05$) were retained. The resulting regression model is:

$$\text{granule size } (\mu\text{m}) = 536.2 - 326.1A - 184.6T + 226.5S + 30.60H + 164.4A^2 + 145.4T^2 + 123.3AT - 110.7AS \quad (2)$$

where A is the scaled airflow rate, T is the scaled inlet air temperature, S is the scaled spray rate and H is the scaled inlet air humidity.

In order to evaluate the adequacy of the empirical model, an analysis of the residuals was performed (Montgomery, 1991). Figure 2 shows that the residuals were normally distributed. It can therefore be concluded that the model proposed in equation (1) fitted the observed granule size adequately.

The model based on the completed data differs to some extent from the model based on the incomplete data (Rambali et al, 2001b). The inlet airflow rate becomes important in the new model; it has significant interaction effects with the inlet air temperature and the spray rate. Due to the large estimated granule size for the missing runs, the quadratic effects of these variables became significant. Also, in the new model the deformation effects of the inlet airflow rate are included (lower airflow resulted in larger granules), which was not the case in the old model.

Figure 3a shows the residuals calculated from multiple regression models based on the incomplete data (Rambali et al., 2001b) and on the completed data, obtained via the deepest regression method (equation 1). The variation of the residuals in both models is comparable. The residuals are not correlated with each other, indicating that both models are different from each other. This difference could be related to the effect of the airflow rate.

The contour plots based on equation (1) are given in figure 4. These contour plots are comparable with the contour plots based on the incomplete data (Rambali et al., 2001b), except at high airflow rate. By increasing the airflow rate at constant fluid bed process settings, it is expected that the granule size decreases. However, based on the contour plots

(especially figures 4g, h, I) a quadratic effect of the airflow rate is observed, which means that depending on the spray rate settings, the granule size is larger at the high level of the airflow rate than at the central level. It was noted that at the high airflow level (runs 11, 18, 27) the size of the granules showed a larger variation than at other runs. In those runs relatively large amounts of large granules occurred together with relatively large amounts of fines. Probably, at high airflow rate the fluidization profile is different, resulting in a inhomogeneously wetted powder bed. Parikh (1997) has described that fluidization of the powder bed depends on several factors, including airflow rate. More research is needed to investigate the effect of the process parameters on the fluidization and on the granule size.

The contour plots confirmed that the granule size obtained at low settings of the spray rate and central level of the inlet airflow rate and inlet air temperature (figure 4 d, e, g and h) was acceptable (between 300 μm and 500 μm).

Two additional runs were performed in order to evaluate the models (table 3). The observed granule sizes were within the predicted confidence interval of each model and therefore the proposed models were valid. The granule size predicted by the models based on the “complete data” were closer to the observed granule size than that of the granule sizes predicted by the models based on the incomplete data.

3.3 Fundamental model

In the previous study (Rambali et al., 2001a) a quadratic model for the granule size was proposed based on fundamental granule growth variables (powder bed moisture after the spraying cycle (M) and a measure for the droplet size (R)) and the deformation variable (A) (force exercised by the airflow rate).

Assuming a quadratic model for the fundamental model, the granule size for the missing runs was estimated by the deepest regression method on the proposed model.

The estimated granule size of the overwetted runs in the fundamental model is shown in table 2. Note that the estimated granule size of the overwetted runs is larger than the granule size of run 12 except for run 26. However, the estimated granule size of run 26 is still quite large, indicating that at these parameter process settings the granulation is unfavorable.

In both models (empirical and fundamental), the largest granule size was estimated for the most unfavorable process settings (runs 5 and 24). Compared with the estimated granule size by the deepest regression method in the empirical model, the estimated granule size in the fundamental model is smaller for some runs (5, 24, 25, 26) and larger for others (1, 6, 20, 21). The difference between the two models is due to difference between the proposed models.

Based on the completed data, a stepwise regression (hierarchical) was applied, which means that non-significant parameters were sequentially eliminated and only significant coefficients ($P < 0.05$) were retained. The resulting regression model is:

$$\text{Granule size } (\mu\text{m}) = 565.5 + 222.3M + 190.7R - 189.0A + 121.9M^2 + 170.21A^2 - 105.0AR \quad (2)$$

where A is the scaled airflow rate, M is the scaled powder bed moisture content and R is the scaled measure for the droplet size.

The regression model based on the the completed data (eq. 2) shows that the airflow rate again is found to be important, as in the empirical model. This model contains different significant interaction effects than in the model based on the incomplete data (Rambali et al, 2001b). In order to compare the fundamental models, based on the complete and the incomplete data

respectively, the residuals of both models were compared to each other. The residuals were highly correlated with each other, indicating that both models are comparable to each other (figure 3b). Although both model consists of different interaction effects, the residuals were comparable, which indicates that probably the fundamental model is more robust to censored data compared with the empirical model. The residual analysis of the fundamental model based on the completed data shows, except for residual of run 15, the residuals are normally distributed (figure 5). The residual of run 15 was also large in the other granule size models based either on the completed data or incomplete data. This means that the predicted models are biased or that the result of this run is due to experimental error. Run 10 with the same process settings and comparable powder bed moisture amount, except for the inlet air humidity, resulted in significantly larger granule size. As the effect of the inlet air humidity on the granule size is not large when compared with the other process variables (see contour plots of figure 4), it seems that the result of run 15 is contributed by experimental error. However, due to planning run 15 was not replicated for confirmation of the experimental error cause.

The contour plots based on equation (2) are given in figure 6 for the fundamental model. These contour plots are comparable with the contour plots of the regression model based on incomplete data, in the sense that acceptable granule sizes were obtained at central and high inlet airflow. However, the same deviation is observed at high inlet airflow rate as in the contour plots of the empirical model (figure 4). At high inlet airflow level ($1100 \text{ Nm}^3/\text{h}$) and at powder bed moisture amount settings of $\pm 33\%$ (wt/wt) smaller granules were expected than at central inlet airflow level ($800 \text{ Nm}^3/\text{h}$) and at the powder bed moisture amount settings. As mentioned this result could be explained by the powderbed fluidization at high level of the airflow.

Two additional runs were performed in order to evaluate the models (table 3). The observed granule size was within the predicted confidence interval of each model and therefore the proposed models were considered valid. The predicted granule sizes with the model based on the “complete data” were comparable with the predicted granule size based on the incomplete data.

4.0 Conclusions

The estimated granule size of the unsuccessful runs by the deepest regression method was acceptable (larger than 811 μm) in almost all cases. Therefore, the deepest regression method seems to be useful for estimating missing data in experimental design. The models based on the incomplete data and “complete data” were comparable in the sense that the contourplots based on the respective models and the predicted granule size were comparable. Some interesting effects of the inlet airflow were observed at the overwetted state of the powder bed in the models based on the “complete data”. It seems that in the overwetted powder bed, the airflow rate becomes important, due to different fluidization. Probably due to inhomogen fluidization, larger granules are obtained at high level of airflow compared with central level of it. The effect of the inlet airflow on the overwetted powderbed needs further study.

References

- Bai, Z. and He, X., 1999. Asymptotic distributions of the maximal depth estimators for regression and multivariate location. *Ann. Statist.*, 27, 1616-1637.
- Doornbos, D.A., 1981. Optimisation in pharmaceutical sciences. *Pharm. Weekblad Sc. Ed.*, 3, 539-577.
- Dussert, A., Chulia, D., Jeannin, C., Ozil, P., 1995. Parametric Study of fluidized-bed granulation of a low density micronized powder. *Drug Dev. Ind. Pharm.*, 21, 1439-1452.
- Gordon, M.S., 1994. Process considerations in reducing tablet friability and their effect on in vitro dissolution. *Drug Dev. Ind. Pharm.*, 20, 11-29.
- Gorodnichev, V.I., El-Banna, H.M., Andreev, B.V., 1981. The Construction of uses of factorial designs in the preparation of solid dosage forms, part 2: granulation in a fluidized bed. *Pharmazie*, 36, 270-273.
- Lipps, M., Sakr, A.M., 1994. Characterization of wet granulation process parameters using response surface methodology., 1. top-spray fluidized bed. *J. Pharm. Sc.*, 83, 937-947.
- Merkku, P., Lindqvist, A-S., Leiviska, K., Yliruusi, J., 1994. Influence of granulation and compression process variables on flow rate of granules and on tablet properties, with special reference to weight variation. *Int. J. Pharm.*, 102, 117-125.
- Meshali, M., El-Banna, H.M., El-Sabbagh, H., 1983. Use of a fractional factorial design to evaluate granulations prepared in a fluidized bed. *Pharmazie*, 38, 323-325.
- Miyamoto, Y., Ogawa, S., Miyajima, M., Sato, H., Takayama, K., Nagai, T., 1995. An evaluation of process variables in wet granulation. *Drug Dev. Ind. Pharm.*, 21, 2213-2225.
- Montgomery, D.C., 1991. Design and analysis of experiments, John Wiley & Sons, Singapore.

- Parikh, D.M., Bonck, J.A., Mogavero, M., 1997. Batch fluid bed granulation. In Parikh, D.M. (Ed), Handbook of pharmaceutical granulation technology, Marcel Dekkers, INC., New York, pp. 227-302.
- Rambali, B., Baert, L., Thoné, D., Massart, D.L., 2001a. Using experimental design to optimise the granulation process in fluid bed. *Drug Dev. Ind. Pharm.*, 27, 53-61.
- Rambali, B., Baert, L., Massart, D.L., 2001b. Using experimental design to optimize the process parameters in fluidized bed granulation on a semi-full scale. *Int. J. Pharm.*, 220, 149-160.
- Rousseeuw, P.J. and Hubert, M., 1999. Regression depth. *Journal of the American Statistical Association*, 94, 388-433.
- Rousseeuw, P.J., Van Aelst, S., Rambali, B., and Smeyers-Verbeke, J., in press. Deepest regression in analytical chemistry. *Analytica Chimica Acta*, 446, 245-256.
- Schaafsma, S., 2000. Down-scaling of a fluidised bed agglomeration process. PhD dissertation, Rijksuniversiteit Groningen, 159 pp.
- Van Aelst, S., and Rousseeuw, P.J. 2000., Robustness of deepest regression. *J. Multivariate Anal.*, 73, 82-106.
- Van Aelst, S., Rousseeuw, P.J., Hubert, M., and Struyf, A. 2002., The deepest regression method. *J. Multivariate Anal.*, 81, 138-166.
- Vojnovic, D., Moneghini, M., Rubessa, F., 1995. Experimental design for a granulation process with “a priori” criterias. *Drug Dev. Ind. Pharm.*, 21, 823-831.

Table 1: Process parameters with their settings
in the face centered central composite design

process parameter	level		
	low	central	high
inlet air flow rate (Nm ³ /h)	500	800	1100
inlet air temperature (°C)	40	55	70
spray rate (g/min)	240	290	340
inlet air humidity (g/kg)	6	10	14

Table 2: Results of the face-centered central composite design

run	process parameters					responses	predicted granule size ^d (μm)	
	inlet air flow rate (Nm ³ /h)	inlet air temperature (°C)	spray rate (g/min)	inlet air humidity (g/kg)	powderbed moisture content (%)	granule size (μm)	empirical model	fundamental model
1	500	70	340	6,01	28,6	>811 ^b	1236	1321
2	1100	40	340	6,01	27,0	592		
3	1100	40	240	6,01	13,6	432		
4	800	55	290	10,03	20,3	497 ^a		
5	500	40	340	14	47,6	>811 ^c	1801	1505
6	500	40	240	6,01	38,0	>811 ^b	1153	854
7	800	55	290	10,03	20,3	484 ^a		
8	800	55	290	14	22,7	572		
9	1100	70	240	6,01	0,0	336		
10	500	70	240	14	16,1	584		
11	1100	70	340	6,01	0,0	518		
12	1100	40	340	12,23	33,4	811		
13	800	55	290	10,03	20,3	493 ^a		
14	1100	40	240	12,23	22,6	429		
15	500	70	240	6,01	15,7	421		
16	1100	70	240	12,23	0,0	342		
17	800	70	290	10,03	4,6	478		
18	1100	70	340	12,23	0,0	610		
19	800	55	290	6,01	17,2	533		
20	500	70	340	14	28,9	>811 ^b	1236	1324
21	500	55	290	10,03	34,5	>811 ^b	1018	1099
22	800	55	290	10,03	20,3	585 ^a		
23	800	55	340	10,03	26,1	736		
24	500	40	340	6,01	44,3	>811 ^c	1801	1473
25	500	40	240	14	42,7	>811 ^b	1153	900
26	800	40	290	10,03	35,1	>811 ^b	911	702
27	1100	55	290	10,03	6,2	409		
28	800	55	240	10,03	12,2	414		
29	800	55	290	10,03	20,3	537 ^a		
30	800	55	290	10,03	20,3	487 ^a		

a = replicates

b = overwetted

c = not performed

d = predicted granule size by depth regression method

Table 3: granule properties of the optimal and variable runs

		observed granule size (μm)	expected granule size (μm)			
			empirical model		fundamental model	
			c	d	c	d
Optimal ^a	run 1	363	448 ± 111	426 ± 121	427 ± 149	425 ± 142
	run 2	445				
Variable run ^b		394	375 ± 116	384 ± 122	412 ± 140	413 ± 140

a = air temperature (55°C), airflow rate (885 Nm³/h), spray rate (290 g/min) and inlet airhumidity

b = air temperature (70°C), airflow rate (950 Nm³/h), spray rate (290 g/min) and inlet airhumidity

c = predicted granule size by the regression model based on the "incomplete data"

d = predicted granule size from the regression model based on the "complete data"

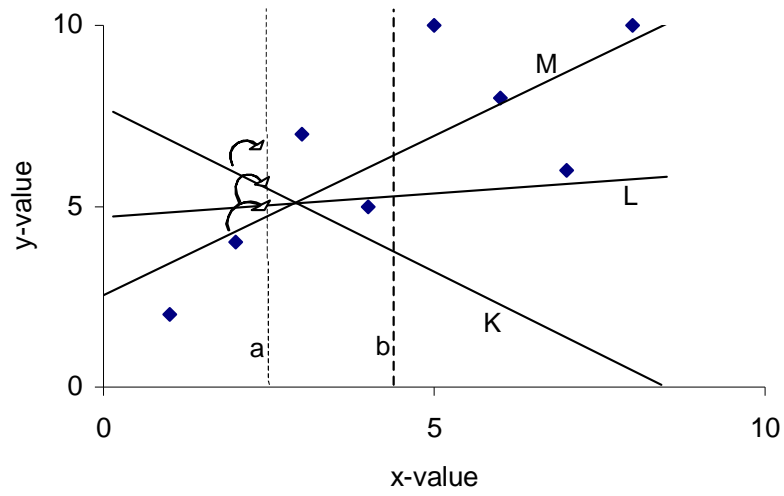


Figure 1: data set with three models, K, L and M with depth regressions of respectively 0, 1 and 2.

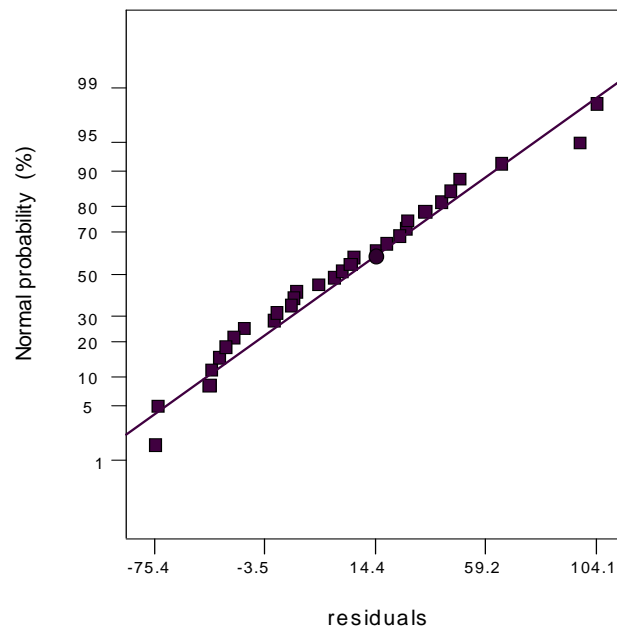


Figure 2: Residuals in function of the expected normal, based on the empirical model which was obtained from deepest regression modelling.

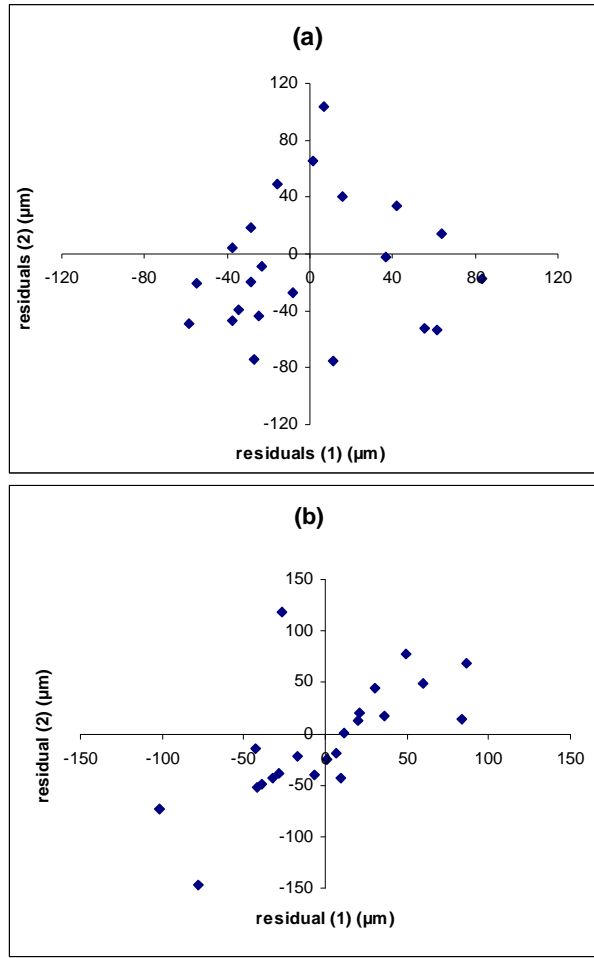


Figure 3: Residuals (a) from the regression models based on the incomplete data and residuals (b) from the regression models ((a) empirical model and (b) fundamental model)) based on the “complete data”.

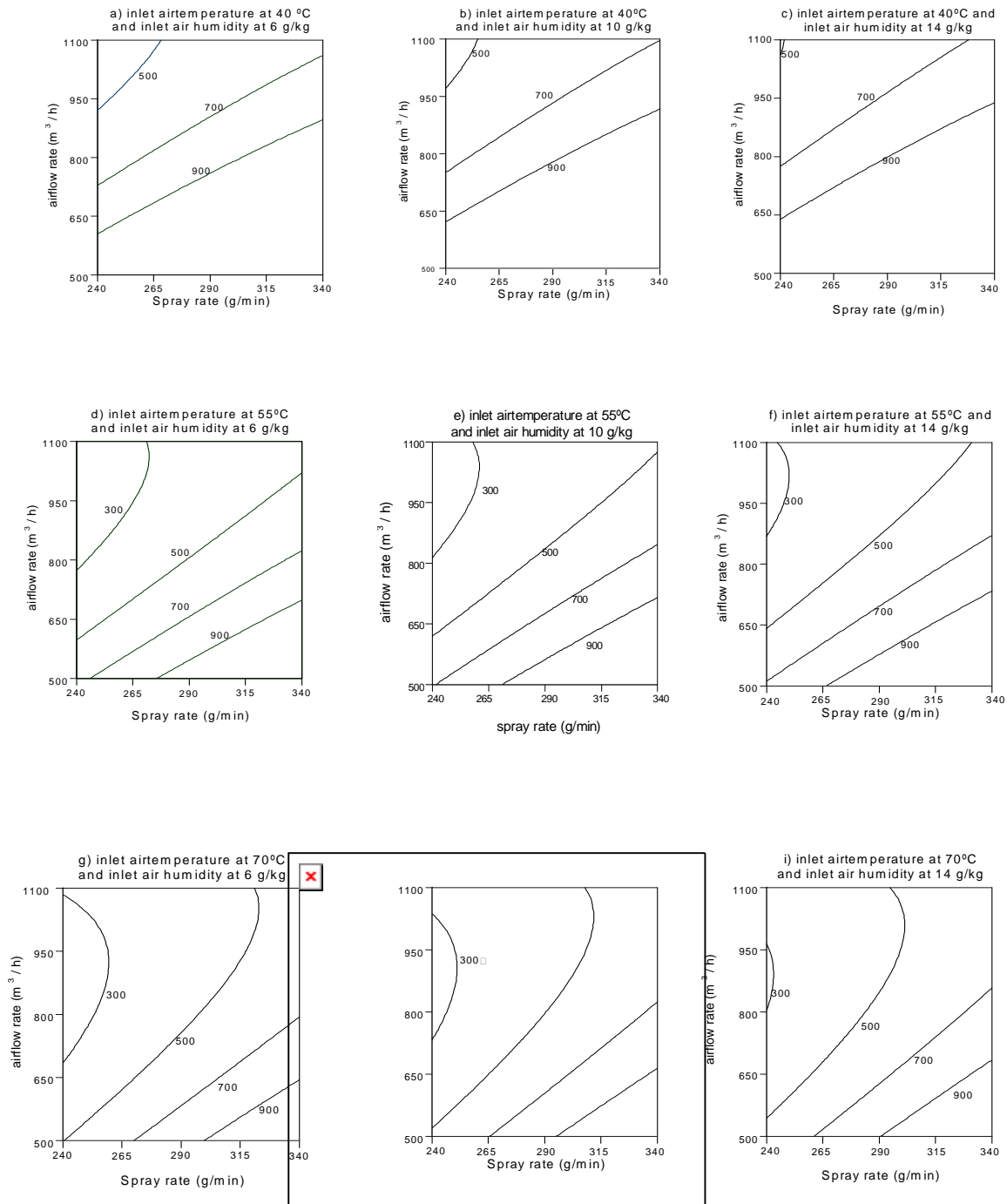


Figure 4: Contour plots of the predicted granule size based on the empirical model of the process variables.

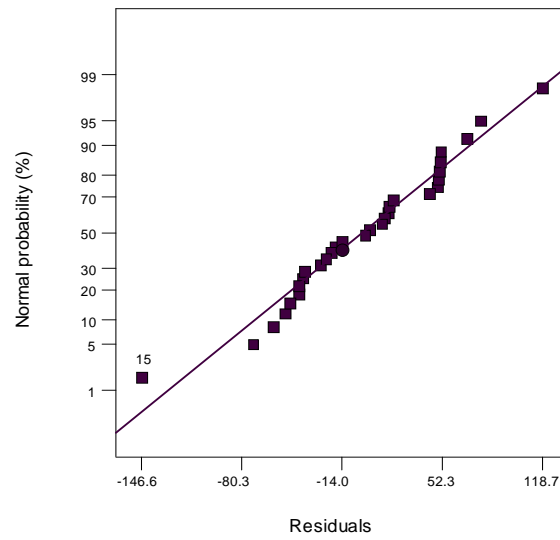


Figure 5: Normal probability plot of the residuals obtained from the fundamental model for the granule size.

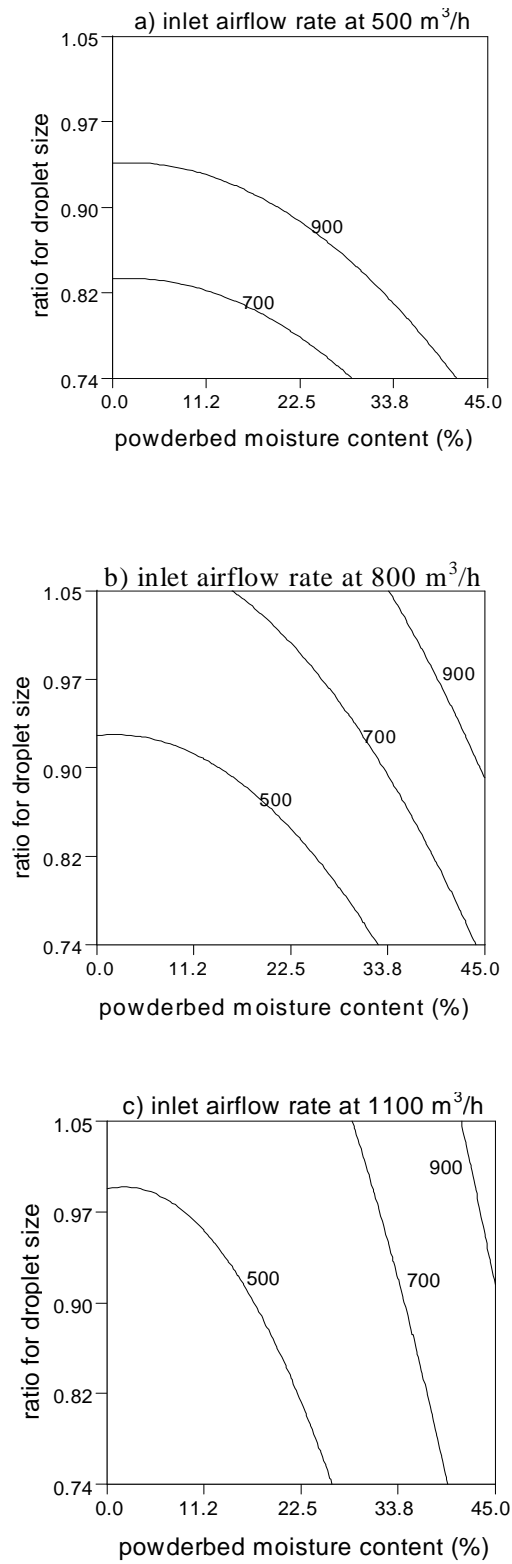


Figure 6: Contour plots of the granule size predicted from the fundamental model based on the “complete data”.